Genetic code expansion for multiprotein complex engineering

Genetic code expansion for multiprotein complex engineering

Christine Koehler[1], Paul Felix Sauter[2], Mirella Wawryszyn[2], Gemma Estrada Girona[1], Kapil Gupta[4], Jonathan J. M. Landry[1], Markus Hsi-Yang Fritz[1], Ksenija Radic[1], Jan-Erik Hoffmann[1], Zhuo Angel Chen[8], Juan Zou[8], Piau Siong Tan[1], Bence Galik[6], Sini Junttila[6], Peggy Stolt-Bergner[6], Giancarlo Pruneri[7], Attila Gyenesei[6], Carsten Schultz[1], Moritz Bosse Biskup[2], Hueseyin Besir[1], Vladimir Benes[1], Juri Rappsilber[8,9], Martin Jechlinger[1], Jan O. Korbel[1], Imre Berger[4,5], Stefan Braese[2,3], Edward A. Lemke[1]*

[1] European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany

[2] Karlsruhe Institute of Technology (KIT), Institute of Organic Chemistry, Fritz-Haber-Weg 6, 76131 Karlsruhe

[3] Karlsruhe Institute of Technology (KIT), Institut für Toxikologie und Genetik, Hermann-von-Helmholtz Platz 1, Campus Nord, 76344 Eggenstein-Leopoldshafen

[4] European Molecular Biology Laboratory (EMBL), Grenoble, 71 avenue des Martyrs, CS 90181, 38042 Grenoble Cedex 9, France

[5] The School of Biochemistry, University of Bristol, Bristol BS8 1TD, United Kingdom

[6] Vienna Biocenter Core Facilities (VBCF GmbH), Dr. Bohr-Gasse 3, A-1030 Vienna, Austria

[7] Division of Pathology and Laboratory Medicine, European Institute of Oncology, Milan, Italy

[8] Wellcome Trust Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK

[9] Chair of Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany

Correspondence to lemke@embl.de

Abstract

We present a protein engineering tool that enables site-specific introduction of unique functionalities in a recombinantly produced eukaryotic protein complex. We demonstrate the versatility of this efficient and robust protein production platform “MultiBacTAG” i) to fluorescently label target proteins and biologics using click chemistries, ii) for glycoengineering of antibodies, and iii) for structure–function studies of novel eukaryotic complexes using single molecule FRET as well as site-specific cross-linking strategies.

Main text

The generation of sufficient quantities of eukaryotic protein complexes is frequently the first and limiting step for the study of molecular mechanisms using numerous biophysical and biochemical assays. Furthermore, expression of many eukaryotic proteins or protein complexes at scales relevant for biotechnological or pharmaceutical purposes, such as biologics, is frequently a daunting task. *Escherichia*

coli is one of the most popular organisms for recombinant protein production, but many proteins and in particular eukaryotic protein complexes cannot be expressed in such simple organisms. Over the last decade, the so-called MultiBac system has established itself among the most widely used systems in basic and applied research on eukaryotic protein complexes production [1, 2]. A particularly attractive feature of MultiBac is the ability to rapidly shuffle proteins, introduce mutations and generate diverse complexes in a user-friendly format to achieve high-yielding expression in insect cell lines derived from *Spodoptera frugiperda* (Sf) or *Trichoplusia Ni* [3]. The power and versatility of this platform could be dramatically enhanced by providing the means to site-specifically engineer diverse custom functionalities into protein complexes.

Genetic code expansion (GCE) is arguably one of the most potent protein engineering technologies, as it allows noncanonical amino acids (ncAAs) harbouring unique functionalities to be encoded site-specifically into a protein of interest (POI). This method has been furthest developed in *E. coli*, in which more than 200 different ncAAs can be introduced anywhere in a polypeptide chain by simply introducing a rare codon (typically the Amber TAG stop codon) in the coding gene of the POI (for reviews, see ref. 4-6). The POI^{TAG} is expressed in an organism that harbours an additional orthogonal tRNA/tRNA-synthetase pair (tRNA/RS), in which the enzyme active site is commonly modified to recognize only a specific ncAA. As such, the Amber codon is repurposed as a sense codon only when the ncAA is present in the growth medium.

We set out to implement the GCE system in MultiBac/insect cells in order to combine advanced protein engineering techniques with convenient, high-yielding recombinant eukaryotic protein complex generation. We chose to work with the pyrrolysine $tRNA^{Pyl}$/PylRS from *Methanosarcina mazei*, as it has already been transferred to a variety of eukaryotic organisms including animals and because most of the available ncAAs have meanwhile been encoded by this system [4-6].

MultiBac consists of one acceptor and several donor plasmid modules that access a baculoviral genome optimized for multigene expression (**Fig. 1**)[3]. The test system consisted of plasmids encoding the wild-type (WT) PylRS from *M. mazei*, a gene cassette for the cognate Amber suppressor tRNA and a reporter protein, mCherry-$GFP^{39\rightarrow TAG}$. The ratio of GFP signal to mCherry provides a convenient readout of the efficiency of Amber suppression as detected by flow cytometry (FC). Subsequently, the system can be tested by transient transfection of Sf21 cells or used to generate a multigene fusion plasmid following established protocols (**Supplementary Fig. 1**, **Supplementary Note 1**)[3]. We utilized the modularity of the MultiBac system to test various known tRNA expression cassettes driven by external U6 PolIII promoters which were used before for successful GCE in other eukaryotes including mammalian cell cultures [7-9] and *D. melanogaster* [10, 11]. As PolIII promoter were not documented for Sf21, a tRNA cassette using U6 promoter from *Bombyx mori* [12], an insect species closely related to *S. frugiperda*, was also tested. Surprisingly, and despite critical external PolIII elements largely considered to be conserved across species (for a comparison of snRNA U6 genes across species see **Supplementary Fig. 2**), no reporter POI expression was detected in any of those cases (**Supplementary Fig. 3**).

Therefore, to identify a potentially useful promoter, we resorted to sequencing and annotating the genome of Sf21 cells (**Supplementary Note 2, Supplementary Table 1**). We identified eight snRNA U6

genes and a dicistronic tRNA expression cassette with a gene architecture analogous to that previously used for efficient GCE in *S. cerevisiae* (**Supplementary Fig. 4, 5**)[13]. As identified by FC analysis, only six U6 driven tRNA constructs allowed for efficient Amber suppression (**Supplementary Fig. 5, 6**).

Choosing U6 promoter 2, we generated a new MultiBac baculoviral genome in which the $tRNA^{Pyl}$/PylRS pair was directly integrated into the viral backbone at the Cre/loxP site (**Fig. 1**, **Supplementary Fig. 1**), termed MultiBacTAG (superscript ^{WT} or ^{AF} for two different PylRS mutants enabling incorporation of different ncAAs shown in **Fig. 1**)[14-16]. The resulting Baculovirus maintains the advantageous features of the MultiBac/insect cell system, including modularity, protease deficiency and delayed insect cell lysis [3] (further details in **Supplementary Fig. 1**).

Figure 2 summarizes an expression test using different reporters and ncAAs. Gratifyingly, expression of the bulky ncAA cyclooctyne-lysine (SCO) using $MultiBacTAG^{AF}$ yielded approximately 2 mg of $GFP^{39\rightarrow SCO}$ (**Fig. 2a)** from a 1 L culture, which is only five fold lower than the average yield of this simple reporter in state of the art *E. coli* GCE systems for the same tRNA/RS and ncAA [14-16] (**Supplementary Fig. 7** for mass spectrometry (MS) validation, **Supplementary Fig. 8** for full-size SDS-PAGE and **Supplementary Table 2** for an overview and comparison of all expression yields in this study). Complementary, the corresponding FC analysis of mCherry-$GFP^{39\rightarrow TAG}$ is shown in **Figure 2b** indicating a ncAA dependent very high efficiency of the GCE MultiBacTAG system (**Supplementary Fig. 8** for complete FC analysis).

MultiBacTAG was further used to engineer Herceptin, a monoclonal antibody and major protein biologic against breast cancer that selectively associates with cancer cells overexpressing the Her2 tumor marker (**Fig. 2** and **Supplementary Fig. 8**)[17]. Amber mutants (A121TAG and A132TAG) were introduced into known permissive sites of the heavy chain of Herceptin [18], and the light and heavy chain were inserted into $MultiBacTAG^{WT\&AF}$. Herceptin was produced intracellularly containing different ncAAs that permit further bioconjugation "click" reactions with diverse substrates ranging from fluorescent dyes to novel glycosyl groups to underline the potential for glycoengineering (**Fig. 2c-f**, **Supplementary Fig. 8-10**, **Supplementary Table 2** for analytics and yields, **Supplementary Note 3** for details on glycan used). In particular *trans*-cyclooctyne-lysine derivatives (TCO*) can undergo particularly fast strain-promoted Diels–Alder [3+2] cycloadditions with tetrazines (SPDAC) and thus allow for exceptionally mild labeling conditions [14-16]. Indeed, TAMRA tetrazine labeled $Herceptin^{121\rightarrow TCO^*\rightarrow TAMRA}$ showed a characteristic positive staining pattern of paraffin embedded human patient samples (**Fig. 2g**, **h**, **Supplementary Fig. 11**, **Supplementary Table 3** for tumor characteristics and HistoIDs).

Next, we utilized the power of the MultiBacTAG system in insect cells to discover novel, hitherto unidentified protein complex dynamics. Genetic and biochemical data suggested the existence of a pentameric transcription factor complex formed between the human TATA-box binding protein (TBP), cognate DNA containing a TATA-box, the general transcription factor TFIIA, and the histone-fold-containing TBP-associated factors TAF11 and TAF13, which constitute a histone-fold pair [19, 20]. We used MultiBacTAG to modify TAF13 in a co-expression experiment with WT TAF11 by using a dual expression cassette inserted into MultiBacTAG virus. Single molecule (sm) Förster Resonance Energy Transfer (FRET) has emerged as a powerful tool to measure distances in proteins between a site-specifically installed donor and acceptor dye pair [21]. We generated a $TAF13^{20\rightarrow SCO}$ mutant and labeled this in a SPDAC

reaction with a sm suitable tetrazine derivative of the donor dye Alexa488. We also labeled a reactive cysteine in TAF13$^{20\rightarrow SCO}$ with a maleimide derivative acceptor dye Alexa 594 (detailed in **Supplementary Fig. 12)**. We then performed smFRET measurements of the TAF11-TAF13$^{20\rightarrow A488,\ 37\rightarrow A594}$ complex. As shown in **Figure 3a**, we detected a population at E_{FRET}= 0.8, which can provide an important distance constrain for further structural model building.

To directly probe protein-protein binding, we designed another mutant that we speculated to be located at binding interfaces. We inserted the ncAA DiAzKs (**Fig. 3b**, **Supplementary Note 4** for synthesis of DiAzKs), which harbours an efficient diazirine protein cross-linker [7, 22] to generate a TAF11/TAF13$^{34\rightarrow DiAzKs}$ complex. We then performed a set of photo-cross-linking experiments followed with subsequent SDS-PAGE and Western Blot (WB) analysis, as summarized in **Figure 3c** (detailed in **Supplementary Fig. 13**). While TAF11/TAF13$^{34\rightarrow DiAzKs}$ yielded a single band cross-link product, a double band appeared in a TBP dependent fashion after UV excitation. SDS PAGE and WB analysis showed that none of the double-bands contained TBP, but had an electrophoretic mobility expected for the TAF11/TAF13 complex. As this indicates a conformational change induced by TBP, we used cross-linking/MS to reveal the actual residues involved. As shown in **Supplementary Fig. 14** and **Supplementary Table 4** we detected five regions of TAF11 to link with TAF13$^{34\rightarrow DiAzKs}$. One region, TAF11$^{146\text{-}149}$ showed marked reduction in linkage in the presence of TBP (Mann Whitney U test, *p<0.05* in both biological replica) (**Fig. 3d, e**). In contrast, cross-links in region TAF11$^{151\text{-}155}$ shown in **Figure 3d**, stayed largely unaffected, indicating that TBP induces specific conformational dynamics at the interface to the TAF11$^{146\text{-}149}$ region, when a TAF11/TAF13/TBP complex is formed (a trimeric complex was also confirmed using size exclusion chromatography **Supplementary Fig. 15**). Our results hint at different modes of assembly involving TAF11, TAF13 and TBP in the absence of cognate DNA and TFIIA (**Fig. 3f**), and set the stage to structure-function determination of the TAF11/TAF13/TBP complex in an integrative approach. Such cross-linking studies can provide invaluable information about solution state dynamics and be used to map dynamic regions complementary to data generated by other structural biology approaches.

In summary, we present here a MultiBac-based system for efficient site-specific incorporation of functionalized amino acids into protein complexes by GCE in Baculovirus/insect cells. MultiBacTAG combines the advantages of high-level expression of even very large eukaryotic protein assemblies offered by the MultiBac system, with a means to engineer and analyze these complexes and their interactions. As the components of the GCE system are inserted into the backbone of MultiBac, the system can be applied readily by the user without prior experience or training in GCE, which maintains the user-friendliness of the system, so that existing MultiBac/insect cells users should be able to move their system to MultiBacTAG without encountering many hurdles. We showed here a selection of applications for MultiBacTAG, ranging from fluorescence labeling of specific targets, to engineering therapeutic protein biologics compatible with human tissue studies and glycoengineering. Engineering of monoclonal antibodies is a contemporary challenge as part of improving pharmaceuticals where high batch-to-batch reproducibility and site-specific chemical modifications are needed, which is a demand that MultiBacTAG combined with click-chemistry intrinsically fulfils. In addition, we used MultiBacTAG to study the formation and conformational dynamics of multicomponent transcription factor complexes using smFRET and site-specific cross-linking. Despite our yields and levels of Amber suppression

efficiency already being satisfying, the 99.6% completed genome of Sf21 presented in this work, will facilitate further genetic engineering of this cell line for protein production using GCE, as e.g. release factor or tRNA expression tuning [4-6]. We anticipate that MultiBacTAG in insect cells will enable a wide range of possibilities for custom protein design for biotechnology and pharmaceutical applications, and be highly useful in the dissection of protein complexes and their functional interactions by unlocking these biological assemblies. This is made possible only by the power of the chemistry that is enabled by site-specific modification through GCE.

Experimental Procedures

Methods and any associated references are available in the online version of the paper.

Author Contributions

C.K. planned and performed experiments, and co-wrote the manuscript. E.A.L. planned experiments, conceived the project and co-wrote the manuscript. P.F.S., M.W., M.B.B., S.B., G.E.G, J.J.L, M.H-Y.F., B.G., S.J., P.S.B., G.P., A.G., H.B., V.B., J.O.K., K.G., I.B., K.R., M.J., J.E.H, C.S., Z.A.C, J.Z., J.R., P.S.T. provided critical instrumental and analytical expertise or reagents.

Acknowledgements

We thank all members of our laboratories for helpful discussions. E.A.L., C.K., P.S., M.W. and S.B. acknowledge funding from the BW Stiftung. E.A.L. acknowledges additional support from the Emmy Noether program. E.A.L. and C.S. are grateful for funding by SPP1623 of the Deutsche Forschungsgemeinschaft. I.B. is funded by the European Commission Framework Programme 7 (FP7) ComplexINC project (contract nr. 279039). P.S.B. acknowledges funding from the Laura Bassi Centres of Expertise initiative for the Centre of Optimized Structural Studies, project 253275. MW thanks the KSOP for financial support. P.S.T is supported by the EMBL Interdisciplinary Postdoc Programme (EIPOD) under Marie Curie Actions COFUND. The Wellcome Trust generously funded this work through a Senior Research Fellowship to J.R. (103139), a Centre core grant (092076) and an instrument grant (108504). We thank also the members of the EMBL Genomics Core Facility for sample processing and sequencing, as well as the EMBL FACS facility for technical support.

Conflict of interest

The authors declare a competing financial interest: a patent application comprising parts of the MultiBacTAG technology here described has been filed.

References:

1. Bieniossek, C., Imasaki, T., Takagi, Y. & Berger, I. MultiBac: expanding the research toolbox for multiprotein complexes. *Trends in biochemical sciences* **37**, 49-57 (2012).
2. Crepin, T. et al. Polyproteins in structural biology. *Current opinion in structural biology* **32**, 139-146 (2015).
3. Fitzgerald, D.J. et al. Protein complex expression by using multigene baculoviral vectors. *Nature methods* **3**, 1021-1032 (2006).
4. Lemke, E.A. The exploding genetic code. *Chembiochem : a European journal of chemical biology* **15**, 1691-1694 (2014).
5. Liu, C.C. & Schultz, P.G. Adding new chemistries to the genetic code. *Annual review of biochemistry* **79**, 413-444 (2010).
6. Chin, J.W. Expanding and reprogramming the genetic code of cells and animals. *Annual review of biochemistry* **83**, 379-408 (2014).
7. Chatterjee, A., Xiao, H., Bollong, M., Ai, H.W. & Schultz, P.G. Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11803-11808 (2013).
8. Chen, P.R. et al. A facile system for encoding unnatural amino acids in mammalian cells. *Angewandte Chemie* **48**, 4052-4055 (2009).
9. Mukai, T. et al. Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochemical and biophysical research communications* **371**, 818-822 (2008).
10. Bianco, A., Townsley, F.M., Greiss, S., Lang, K. & Chin, J.W. Expanding the genetic code of Drosophila melanogaster. *Nature chemical biology* **8**, 748-750 (2012).
11. Mukai, T., Wakiyama, M., Sakamoto, K. & Yokoyama, S. Genetic encoding of non-natural amino acids in Drosophila melanogaster Schneider 2 cells. *Protein science : a publication of the Protein Society* **19**, 440-448 (2010).
12. Hernandez, G., Jr., Valafar, F. & Stumph, W.E. Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity. *Nucleic acids research* **35**, 21-34 (2007).
13. Hancock, S.M., Uprety, R., Deiters, A. & Chin, J.W. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *Journal of the American Chemical Society* **132**, 14819-14824 (2010).
14. Nikic, I. et al. Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angewandte Chemie* **53**, 2245-2249 (2014).
15. Plass, T. et al. Amino acids for Diels-Alder reactions in living cells. *Angewandte Chemie* **51**, 4166-4170 (2012).
16. Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E.A. Genetically encoded copper-free click chemistry. *Angewandte Chemie* **50**, 3878-3881 (2011).
17. Axup, J.Y. et al. Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 16101-16106 (2012).
18. Xiao, H. et al. Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angewandte Chemie* **52**, 14080-14083 (2013).
19. Kraemer, S.M., Ranallo, R.T., Ogg, R.C. & Stargell, L.A. TFIIA interacts with TFIID via association with TATA-binding protein and TAF40. *Molecular and cellular biology* **21**, 1737-1746 (2001).
20. Robinson, M.M. et al. Mapping and functional characterization of the TAF11 interaction with TFIIA. *Molecular and cellular biology* **25**, 945-957 (2005).

21. Tyagi, S. & Lemke, E.A. Single-molecule FRET and crosslinking studies in structural biology enabled by noncanonical amino acids. *Current opinion in structural biology* **32**, 66-73 (2015).
22. Zhang, M. et al. A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nature chemical biology* **7**, 671-677 (2011).

Figure legends:

Figure 1: Overview of the new MultiBacTAG system

The scheme illustrates an overview of the newly established MultiBacTAG system for the expression of multidomain protein complexes in insect cells with different ncAAs for diverse applciations. Several POIs can be combined using tandem recombineering of several donor and one acceptor plasmid (pIDC, pIDK, pIDS and pACEBac1,2) via Cre/loxP sites (violet sphere, more details given in corresponding **Supplementary Fig. 1**) and then be inserted into the Tn7 site in the Bacmid DNA, which contains the tRNA/PylRS pair. After production of the Baculovirus, insect cells can be transduced and the ncAA of choice will be added. Structures of ncAAs used in this work are shown, propargyl-lysine (**1**, PrK), cyclooctyne-lysine (**2**, SCO), Boc-lysine (**3**, BOC), *trans*-cyclooctene-lysine (**4**, TCO*), BCN-lysine (**5**, BCN) and diaziridine-lysine (**6**, DiAzKs).

Figure 2: Characterization of MultiBacTAG, and diverse click labeling of Herceptin, and detection of human cancer

(**a**) SDS-PAGE after purification of $GFP^{39\rightarrow TAG}$ expressed in Sf21 cells transfected with $MultiBacTAG^{AF}$ grown in the presence (+) and absence (-) of 1 mM SCO (**Supplementary Fig. 8** for full-size gels and other ncAAs). The corresponding FC analysis of mCherry-$GFP^{39\rightarrow TAG}$ is shown in (**b**). Shown experiments reveal a clear ncAA dependent protein production and are representative of at least three independent experiments. **(c)** illustrates different labeling reactions between antibody and dye (green dot) or glycan. From top to bottom: i) copper-catalyzed click labeling reaction between a terminal alkyne and an azide. ii) copper-free strain promoted azide alkyne cycloaddition between BCN and an azide containing glycan structure (see **Supplementary Fig. 10** for experimental data) iii) and iv) different SPDAC reactions. (**d**-**f**) UV scans of different labeling reactions on the left and Coomassie-stained SDS-PAGE gels on the right of each panel (full size gels in **Supplementary Fig. 8**). (**d**) Copper-based click chemistry of $Herceptin^{132\rightarrow PrK}$ with fluorescein-azide. (**e**) SPDAC reaction between $Herceptin^{121\rightarrow SCO}$ with TAMRA-tetrazine (Herceptin WT used as negative control). (**f**) SPDAC reaction between $Herceptin^{121\rightarrow TCO^*}$ and TAMRA-tetrazine. (**g**–**h**) $Herceptin^{121\rightarrow TCO^*\rightarrow TAMRA}$ is suitable to detect cancer cells in human patient samples (n=3 for positive and negative tissue samples shown here and in **Supplementary Fig. 11**). Human tumour sections included Her2+ and Her2- (**g,h**, HistoIDs see **Supplementary Table 3**) samples. Images shown are maximum projections of 35 planes spanning 5 µm total. Blue channel: DAPI, red channel: $Herceptin^{121\rightarrow TCO^*}$ labeled with TAMRA-tetrazine.

Figure 3: Cross-linking of TAF11/TAF13/TBP complex

(**a**) A cartoon of the TAF11/TAF13 complex is shown with labelling sites indicated by a green and a red star (donor and acceptor position), as well as FRET efficiency (E) vs stoichiometry (S) plot revealing a population at E=0.8 (the population around E=0 is due to dye photophysics or limited labelling

efficiencies). (**b**) Cross-linking scheme between two proteins using DiAzKs and UV light. (**c**) Shows a Coomassie-stained SDS-PAGE (top) and the corresponding anti-TAF13 WB of the cross-linking experiment of TAF11/TAF13 complex with increasing TBP (1:1:0 (-), 1:1:0.625 (⊕), 1:1:1.25 (+)). (**d**) MS analysis of gel cross-linked products from (**c**) (analysed bands boxed schematically in black), revealing two cross-link regions in TAF11 with $TAF13^{34\rightarrow DiAzKs}$. Sample A and B (both TAF11+$TAF13^{34\rightarrow DiAzKs}$+TBP) are biological replica each with their own reference of TAF11+$TAF13^{34\rightarrow DiAzKs}$ without TBP. Relative abundance of cross-links in presence of TBP were calculated against a reference of TAF11+$TAF13^{34\rightarrow DiAzKs}$ in absence of TBP. To show the variance in the measurements, also the reference was replicated (sample A^0). Center values are the median, error bars show standard deviations based on multiple cross-linked peptides and "n" indicates the number of quantified cross-linked peptides (**Supplementary Fig. 14, Supplementary Table 4** for additional details). (**e**) Annotated high-resolution fragmentation mass spectrum of cross-linked peptide RSAFPK - $FLSK_{DiAz}ELR$, revealing a cross-link of $TAF11^{153}$ to $TAF13^{34}$. A fragment ion annotated with "+P" is a fragment ion that contains the cross-linked partner peptide. "P+P" refers to the intact precursor ion. **f)** TAF13 (blue) and TAF11 (yellow) form a tight complex (top) yielding two cross-links (red). Binding to TBP (shown in grey) results in a trimeric complex (bottom) displaying an altered cross-linking pattern (grey dashed arrow). The complex and cross-links are shown in a cartoon representation, with labeled N- and C-termini.

Online methods

Reagents

If not further noticed chemicals were purchased from Sigma. Noncanonical amino acids were prepared in-house, in the case of DiAzKs, otherwise received from Sirius Fine Chemicals (SiChem, Bremen), in case of PrK, SCO, TCO* and BCN (note, now DiAzKs can also be purchased from SiChem). BOC was purchased from IRIS Biotech (Marktredwitz).

Sequencing and analysis of the Sf21 genome

The Sf21 genome was sequenced by Illumina sequencing technology using 3 types of libraries. Two short-insert paired-end libraries (2x104 bp of ~288 bp insert size and 2x36 bp of ~590 bp insert size), two long-insert mate-pair libraries (2x94 bp and 2x101 bp of ~4500 bp insert size) and one TruSeq Synthetic Long-Read library were generated and sequenced. The data obtained with the last library was assembled into long synthetic reads using the TruSeq Long-Read Assembly app v1.1 available on BaseSpace (Illumina Inc.). At first the paired-end reads were corrected and filtered with SGA (version 0.9.43)[23]. The resulting ~87.2e6 read pairs were used as input to perform contig assembly, scaffolding and gap closing using SOAPdenovo2 (version 2.4) [24]. Second, mate-pair reads were processed with FLASH [25] (version 1.2.6) and all overlapping read pairs were discarded. The resulting ~32.4e6 pairs were employed with SOAPdenovo2 for scaffolding and then gap closing of the previous assembly. Third, the 18.3e4 long synthetic reads were used to scaffold the assembly obtained with paired-end and mate-pair sequencing data. All data types were then finally utilized for a final gap closing step (SOAPdenovo2).

Eight U6 snRNA gene could be found (U6-1 – U6-8), using *Bombyx mori* snRNA U6 isoform E gene as query sequence (RefSeq: AY649381.1), with at least 400 bp upstream (promoter region) and 100 bp downstream sequences (termination signal) (**Supplementary Fig. 4**). We decided to work with U6 promoter and the 3'termination signal out of the second scaffold (17011_2962_3036_+), which was found, and called this U6 promoter, U6(Sf21)-2.

Mulitbac Baculovirus system for transduction of insect cells

Construction of amber suppressor genomes (MultiBacTAG^{WT}, MultiBacTAG^{AF}):

We generated a baculoviral genome which contains the genes encoding for both the synthetase and the tRNA for amber suppression by using Cre recombinase mediated insertion into the LoxP present on the MultiBac viral backbone (**Fig. 1**). Thus, the attachment site for Tn7 transposition (mini-attn7) remains fully accessible to accept multigene constructs of target proteins and their complexes. We inserted the expression cassette U6(Sf21)-2-tRNA^{Pyl}-3'term into the pUCDM Donor plasmid module by using ClaI and XbaI restriction enzymes. Next, we added by means of NsiI and XhoI digestion and ligation the MM PylRS or MM PylRS AF into the p10 driven expression cassette, giving rise to MultiBacTAG^{WT} and

MultiBacTAG^{AF} viral genomes, respectively. For all cloning steps of the pUCDM plasmid, BW23474 cells were used to provide the Pir+ background required by the conditional origin present on the Donor [3]. The resulting dual expression plasmid pUCDM-U6(Sf21)-2-tRNA^{Pyl}-3'term-PylRS was transformed into electro-competent DH10MultiBac^{Cre} cells, following established protocols [3,26]. Tetracyclin antibiotic challenge was applied during all transformation steps to ensure maintenance of the pHelper plasmid which encodes for the Tn7 transposase and is required for inserting multigene constructs encoding for target proteins. Cell stocks were validated by preparing composite baculoviral genomes from eight blue colonies each and transfection of Sf21 cells. V_0-virus was harvested after 60 hours of incubation and the V_1-generation was started. Cells were harvested 60 hours after proliferation arrest [3]. Cell pellets were resuspended in 4 x PBS (phosphate-buffered saline) (pH 8), resulting in 1 Mio. cells/ml. Glycerol stocks of cells containing MultiBacTAG^{WT} and MultiBacTAG^{AF} were prepared respectively and from those electrocompetent cells were prepared following standard protocols, and stored at -80 ^{o}C.

Plasmids:

Reporter plasmids:

First a reporter plasmid was constructed. GFP(Y39TAG)-6His and mCherry-GFP(Y39TAG)-6His were separately cloned into Acceptor pACEBacDual plasmid under the polh (Polyhedrin) promoter, using BamHI and PstI restriction enzymes. The resulting pACEBac-Dual-GFP(Y39TAG)-6His and pACEBac-Dual-mCherry-GFP(Y39TAG)-6His acceptors were transformed into cell containing MultiBacTAG^{WT} and MultiBacTAG^{AF}, respectively, for integration into the Tn7 attachment site.

Herceptin:

Synthetic genes encoding for the variable and constant regions of the heavy and light chain of the Herceptin were codon optimized for insect cell expression and inserted into pACEBacDual Acceptor into the polh and p10 driven expression cassettes, respectively. A C-terminal six-histidine tag was fused to the Herceptin heavy chain. Two individual amber mutations were inserted at positions A121 and A132 of the heavy chain.

TAF11/TAF13/TBP complex:

pFastBac-Dual-6HisTAF11/TAF13 was constituted from pFastBac-Dual by inserting the genes encoding for human TBP associated factors 11 (TAF11) and 13 (TAF13) into the polh and p10 driven expression cassettes. TAF11 contains an N-terminal hexa-histidine tag followed by a tobacco etch virus (TEV)-NIa protease site. Two Amber stop codons were introduced separately into the TAF13 gene at positions A20 and K34. Human TATA-Box binding protein (TBP) core (residues 155-333) was cloned into pET28aHis plasmid, resulting in a six-histidine tag at the N-terminal domain of TBP (courtesy of T.J. Richmond, ETH Zurich).

Cell culture

Sf21

Following standard protocols [27], Sf21 cells were cultured in Erlenmeyer flask at 27 °C shaking at 180 rpm, using Sf-900™ III SFM medium at the Protein Expression and Purification core facility (PEPcore) at EMBL, Heidelberg. Cells were split every day to $0.6*10^6$ cells/ml or every third day to $0.3*10^6$ cells/ml. For Bacmid transfection, 3 ml per well of $0.3*10^6$ cells/ml were seeded in a 6-well multidish (Nunclon Delta Surface, Thermo scientific). Bacmid-DNA was prepared and Sf21 cell transfected using FuGENE HD Transfection Reagent (Promega). V_0-virus was harvested after 70 hours post transfection and the V_1-generation started. For small scale test expression, 100 ml of Sf21 cells at $0.6*10^6$ cells/ml were transfected with 0.1 ml of V_1-virus and 1 mM of the respective ncAA was added. As negative control, a 100 ml culture was set up the same way, but without ncAA. After cell proliferation stopped, the cultures were kept another 48-60 hours at 27 °C shaking at 180 rpm. The cells were harvested at 500 rpm for 10 minutes and the pellets were stored at -20 °C.

Flow cytometry analyses

Flow cytometry analyses were done on a BD LSRFORTESSA (BD Biosciences). Therefore Sf21 cells were transduced with the corresponding virus in a 6-well multidish. After three days of incubation time, the cells were harvested at 500 rpm for 10 minutes at 4°C and resuspended in 500 µl sterile 1 x PBS. The suspension was filtered through a cell strainer (Falcon, 70 µm, Fisher scientific) and kept on ice until measurements. Data of 500,000 cells for each sample was acquired and analyzed with FlowJo X software (FlowJo Enterprise).

Protein expression and purification

GFP(Y39TAG) & mCherry-GFP(Y39TAG):

The plasmids pACEBacDual-GFP(Y39TAG)-6His and pACEBac-Dual-mCherry-GFP(Y39TAG)-6His were transformed into cells containing MultiBacTAG (WT and AF variants), and plated on agar plates containing X-Gal and IPTG (for blue/white selection), as well as Ampicillin (100 µg/ml), Kanamycin (30 µg/ml), Tetracycline (10 µg/ml) and Gentamycin (10 µg/ml). Four white colonies each were picked and composite baculoviral DNA prepared. After transfecting Sf21 cells the four V_0-Vvirus preparations were harvested after 60 hours. V_1-virus was produced using all four V_0-viruses in parallel and for each 0.1ml of Virus was added to 100 ml of fresh Sf21 cells. Five cultures were set up in the same way, one for each of the four V_1-viruses, in which ncAA at a final concentration of 1 mM was added and 1 culture without ncAA, as a negative control. After cell propagation stopped, the cells were harvested after additional 48-60 hours.

For purification, cell pellets were resuspended in 4 x PBS (5 mM imidazol, 0.2 mM TCEP, 1mM PMSF) and centrifuged at 40000 rpm at 4 °C using a Beckman ultracentrifuge (SW Ti60 rotor) after sonication. The cleared lysate was incubated on Ni beads for 1-2 hours at 4 °C. The Immobilized metal ion affinity chromatography (IMAC) was carried out by washing with 10 mM imidazol in 4 x PBS (0.2 mM TCEP and 1 mM PMSF), followed by an elution step using 500 mM imidazol in the same buffer. Finally the elution fraction was analyzed by SDS-PAGE and stored at -20 °C.

Herceptin:

For the expression of Herceptin the plasmid pACEBacDual-Herceptin-6His was transformed in both, MultiBacTAG^{WT} and DH10MultiBacTAG^{AF} containing cells. Expression and purification was carried out following the same steps as described above for GFP(Y39TAG).

TAF11/TAF13 complex:

For producing TAF11/TAF13 complex, MultiBacTAG^{AF} was used, for both wild-type TAF11/TAF13 complex, as well as for the amber mutants (see above). Again, the same protocol was followed as described above for GFP(Y39TAG).

The cell pellet was resuspended in 150 ml Tris buffer (25 mM Tris, 150 mM NaCl, 5 mM imidazol, 1 mM PMSF, pH 8) per 1 liter expression culture. After sonication, the insoluble fraction was spin down at 40000 rpm at 4 °C (Beckman SWTi60 rotor). The supernatant was incubated on Nickel beads for 1-2 hours and the protein was eluted after several washing steps with increasing imidazol concentrations. To finalize the IMAC purification procedure, the protein was further purified by size exclusion chromatography (SEC) using a Superdex column, equilibrated before hand with Superdex running buffer (25 mM Tris, 300 mM NaCl, 1 mM EDTA, 1mM DTT, pH 8) and analyzed by SDS-PAGE.

TATA-Box binding protein (TBP), residues 155-333:

pET28aHis-TBP was transformed into BL21(DE3) Rosetta cells and expressed in LB medium at 18°C over night. Cells were harvested by centrifugation (4500 rpm, 20 min., 4 °C) and stored at -20°C.

The cells of 1 liter expression culture were lysed in 20 ml TBP lysis buffer (25 mM Tris, 1 M NaCl, 10 mM imidazol, 1 mM PMSF, pH 8) using a sonicator. After spinning down the insoluble fraction, the cleared supernatant was purified by IMAC. Washing was done with increasing concentration of imidazol and the protein was finally eluted. After loading the protein on a Superdex column, which was equilibrated with Superdex running buffer, the purity was checked by SDS-PAGE analysis.

Single Molecule FRET experiments

Dual labelled TAF11/TAF13$^{20\rightarrow A488,\ 37\rightarrow A594}$ complex were diluted to ~ 100 pM and subject to multiparameter single molecule FRET (smFRET) spectroscopy on a custom built confocal detection setup as detailed previously [28]. In brief, the sample was excited through a 1.2NA 63x Olympus objective with

alternating LASER pulses from a 485 LDH diode Laser and an 570 nm filtered while light LASER (Koheras). Emission signal was split into green and orange color channels, and detected on photon counting diodes (MPD and APD), directed to Hydraharp (Picoquant) counting electronics and analyzed further using IgorPro (Wavemetrics) as detailed previously.[28] The signals intensities were analyzed according to the following equations, with I_A and I_D being the recorded photon counts during donor Laser excitation, and I_A^{dir} the intensity of the acceptor during acceptor LASER excitation. The plot shown in main Figure 4a shows a 2D E_{FRET} vs S plot. At E=0 and S=1 sits the so called "Zero"-Peak which arises from inactive acceptor, and is not of relevance in this analysis. From the known γ (a correction factor for the apparent brightness of our dye pair) and the known R_0 for our dye pair [29], we can estimate that the measured FRET intensity corresponds to an approximate distance (r) of around 30Å.

$$E_{FRET} = \frac{I_A}{\gamma I_D + I_A} = \frac{1}{1 + (\frac{r}{R_0})^6}; S = \frac{I_A + I_D}{I_D + I_A + I_A^{dir}}$$

Cross-linking experiments

Western Blot analysis of cross-linked samples

The cross-linking reactions contained 40 µM of TAF11/TAF13 complex. TBP was added in two different molar ratios to the reaction. The first ratio was 1:1:0.625, TAF11:TAF13:TBP correspondingly. For this ratio, we used 12.5 µM of TBP. The second ratio was 1:1:1.25, which results in 25 µM of TBP per reaction. For each cross-linking experiment, we set up 20 µl reactions containing the respective proteins in Superdex running buffer and incubated the reactions on ice for 2 hours. These reactions were then splitted into 2 x 10 µl, and one of the 10 µl reactions was exposed to UV light. UV irradiation was performed for 15 minutes on ice using a 345 nm filter with an approximately 40 cm distance to the 1000 W lamp. The cross-linking experiments were performed with a TAF13$^{34 \rightarrow \text{DiAzKs}}$ mutant.

For preparing the samples for SDS-PAGE, 5 µl of each reaction was mixed with 35 µl Superdex running buffer and 10 µl 5 x SDS loading dye, then the samples were heated up for 1 minute at 95 °C. 15 µl of these samples were loaded in a well of a 10-well SDS-PAGE (NuPAGE 4-12% Bis-Tris, Thermofisher). After running the gels using MES buffer, they were plotted using the Trans-Blot® Turbo™ Transfer system (Bio-Rad). With the Trans-Blot® Turbo™ Mini Nitrocellulose Transfer Packs (Bio-Rad) the transfer was done in 7 minutes and the membranes were blocked for 1 hour at room temperature with 5% Milk in 1 x PBS. The primary antibodies (anti-TAF13 (Abcam), anti-TBP (kind gift from Laszlo Tora) and anti-Flag (Monoclonal Antibodies Core Facility, EMBL)) was diluted 1:1,000 (for anti-TAF13) and 1:2,000 (for anti-TBP and anti-Flag) in 5% Milk, 1 x PBS and the membrane was incubated over night at 4 °C. After a few washes with 1 x PBS, 0.2% Tween 20, the secondary antibody was incubated for 1 hour at room temperature. For the anti-TAF13 an anti-rabbit secondary antibody (Peroxidase AffiniPure Goat Anti-Rabbit IgG (H+L), Jackson ImmunoResearch) was used in a 1:5,000 dilution in 1 x PBS, 0.2% Tween 20 and for the anti-TBP and anti-Flag antibodies an anti-mouse secondary antibody was diluted 1:10,000 in 1 x PBS, 0.2% Tween 20 (Amersham ECL HRP Conjugated Antibodies, GE Healthcare). After three more

washes with 1 x PBS, 0.2% Tween 20, a chemiluminescence Kit (ECL Western Blot reagent, GE Healthcare) in combination with a Chemidoc Touch system (Biorad) was used to visualize the Western Blot signal.

Sample preparation for mass spectrometric analysis

For mass spectrometric analysis, the cross-linking reaction was set up in the ratio 1:1:1.25, TAF11:TAF13:TBP correspondingly. The TAF13$^{34\rightarrow DiAzKs}$ cross-linked samples were prepared in replicates as given in the text (**Fig. 3**). For each reaction 40 µM of TAF11/TAF13 complex were mixed with 25 µM of TBP in a 30 µl reaction volume, incubated on ice, cross-linked by UV light (15 min, 345 nm filter, 1000 W lamp) and loaded on a SDS-PAGE. 1.5 µl of each reaction were loaded on the same gel in a separate well, which was used to identify the cross-linked species by Western Blot. The gel bands of cross-linked TAF11/TAF13 complexes were excised, in-gel reduced and alkylated, then digested using trypsin following a standard protocol[30]. The peptide mixture was then desalted using C18-Stage-Tips[31] for mass spectrometric analysis.

Mass spectrometric analysis

LC-MS/MS analysis was performed using an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Scientific) applying a "high-high" acquisition strategy. Peptides were separated on a 75 µm x 50 cm PepMap EASY-Spray column (Thermo Scientific) fitted into an EASY-Spray source (Thermo Scientific), operated at 50 °C column temperature. Mobile phase A consisted of water and 0.1% v/v formic acid. Mobile phase B consisted of 80% v/v acetonitrile and 0.1% v/v formic acid. Peptides were loaded at a flow-rate of 0.3 µl/min and eluted at 0.2 µl/min using a linear gradient going from 2% mobile phase B to 4% mobile phase B over 139 minutes, followed by a linear increase from 45% to 95% mobile phase B in eleven minutes. The eluted peptides were directly introduced into the mass spectrometer. MS data were acquired in the data-dependent mode with the top-speed option. For each three-second acquisition cycle, the survey level spectrum was recorded in the Orbitrap with a resolution of 120,000. The ions with a precursor charge state between 3+ and 8+ were isolated and fragmented using high-energy collision dissociation (HCD). Precursor priority for fragmentation was set to "highest charge state" then "most intense". The fragmentation spectra were recorded in the Orbitrap with a resolution of 15,000. Dynamic exclusion was enabled with single repeat count and 60-second exclusion duration.

Identification of cross-linked peptides

The raw mass spectrometric data files were processed into peak lists using MaxQuant version 1.5.3.30[32] with default parameters, except for "FTMS top peaks per 100 Da" was set to 100 and "FTMS de-isotoping" was disabled. The peak lists were searched against the sequences as well as the reversed sequences (decoy) of TAF11 and TAF13$^{34\rightarrow DiAzKs}$ using Xi software (ERI, Edinburgh) for identification of cross-linked peptides and non-cross-linked linear peptides. In the protein sequences, DiAzKs was

represented as “Xd”. Search parameters were as follows: MS accuracy, 6 ppm; MS2 accuracy, 20 ppm; enzyme, trypsin; specificity, fully tryptic; allowed number of missed cleavages, four; fixed modifications, carbamidomethylation on cysteine; variable modifications, oxidation on methionine. The cross-linking reactivity of DiAzKs is towards any other amino acid residues. All fragmentation spectra of all identified cross-linked residue pairs were validated manually. In addition, we identified linear peptides from TAF11, and TAF13. Linear peptides with Xi score above 7 were used for quantitation to estimate the relative protein abundance in each sample.

Quantitation of cross-link data using Pinpoint software

Identified cross-linked peptides and selected linear peptides were quantified based on their MS1 signals. The quantitative proteomics software tool Pinpoint (Thermo Fisher Scientific) was used to retrieve intensities for each cross-linked and linear peptide[33]. To construct the input library of Pinpoint, the sequence of every cross-linked peptide was converted into a linear version with identical mass[34]. The five most abundant signals in the isotope envelope were used for quantitation. The error tolerance for precursor m/z was set to 6 ppm. Signals are only accepted within a window of retention time (defined in the spectral library) ±10 minutes. Manual inspection was carried out to ensure the correct isolation of elution peaks. “Match between runs”[35] was carried out for all cross-linked peptides in Pinpoint interface manually, based on high mass accuracy and reproducible LC retention time.
The signal intensities of cross-linked peptides were normalized against abundance of TAF13, which was calculated as summed signal intensities of seven linear peptides. The relative abundance of cross-links in samples with and without TBP was compared.

Statistsics

QCLMS analysis was repeated in two separated experiments. In experiment I, three samples were analyzed: two $TAF13^{34\rightarrow DiAzKs}$+TAF11 samples (reference and A^0) and one $TAF13^{34\rightarrow DiAzKs}$+TAF11+TBP sample (A). In experiment II, two samples were analyzed: one $TAF13^{34\rightarrow DiAzKs}$+TAF11 sample (reference) and one $TAF13^{34\rightarrow DiAzKs}$ +TAF11+TBP sample (B).

The TAF11 residues that were cross-linked to DiAzKs fall into five regions. For each sample, the relative intensity of cross-links to each region was calculated as the median of all their supporting cross-linked peptides. The numbers of supporting cross-linked peptides (n) for cross-linkages to each TAF11 region were listed in **Figure 3d** and **Supplemental Figure 14**.

Click reactions

Copper-catalyzed alkyne-azide cycloaddition (CuAAC):

Purified protein, which contains an ncAA (Propargyllysine, PrK) with an alkyne group incorporated at the amber stop codon side, was exchanged to 1 x PBS buffer pH 7.5 (0.2 mM TCEP) and 5 nmol were used

for the click reaction, following the protocol as described in ref. [36]. Cycloaddition reactions were followed up by SDS-PAGE.

Strain-promoted alkyne-azide cycloaddition (SPAAC):

Protein, expressed in the presence of 1 mM of BCN (Sichem), was purified and exchanged into 1 x PBS buffer (pH 8). For the labeling reaction 2 nmol of protein mixed with 100 nmol of glycan-azide (PSZ170) were incubated over night at RT [17]. Labeling reactions were loaded on a Superdex column and analyzed by SDS-PAGE.

Strain-promoted Diels-Alder cycloaddition (SPDAC):

Protein, expressed in the presence of 1 mM of SCO (Sichem) or TCO* (Sichem), was purified and exchanged into 1 x PBS buffer (pH 8). For the labeling reaction 1 nmol of protein mixed containing SCO with 5 nmol of TAMRA-Tetrazine (Jena Bioscience) were incubated for 1 hour at RT [16]. In the case of protein harboring TCO*, 5 nmol of protein were used in a reaction with 50 nmol of Tetrazine-5-TAMRA. Labeling reactions were loaded on a Superdex column and analyzed by SDS-PAGE.

Immunofluorescence analysis

Tissue sections were processed for immunofluorescence staining and incubated with Herceptin$^{121\rightarrow TCO*}$ TAMRA labeled antibody (diluted 1:100) overnight, 4 ^{o}C, washed in PBS and mounted in ProLong Gold antifade with DAPI (Invitrogen). Images were obtained on a Leica TCS SP5, LAS AF Version 2.7.3.9723 (Leica Microsystems CMS GmbH). Objective: HCX PL APO lambda blue 63.0 x/1.40 OIL UV.

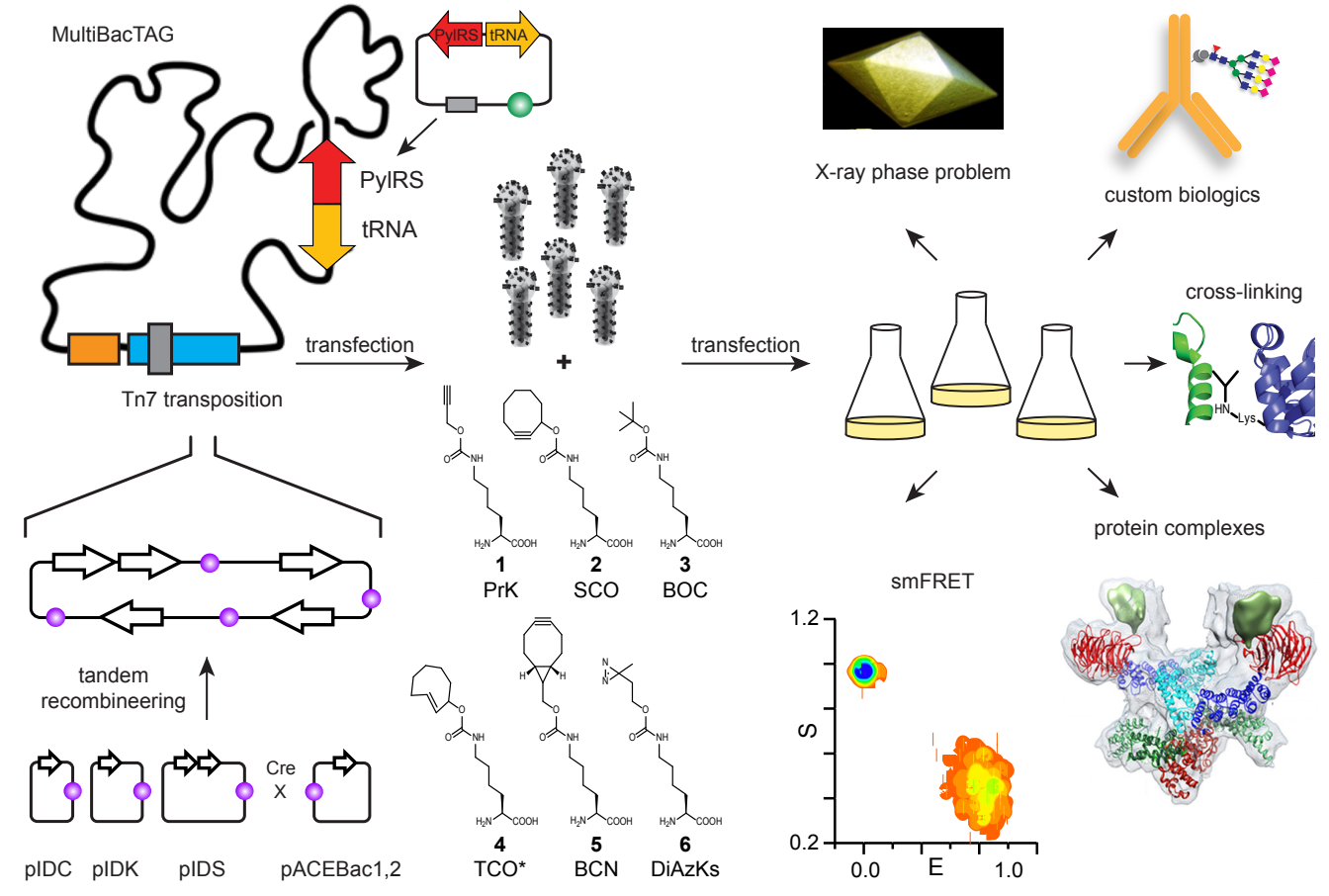
Human Tissue Samples

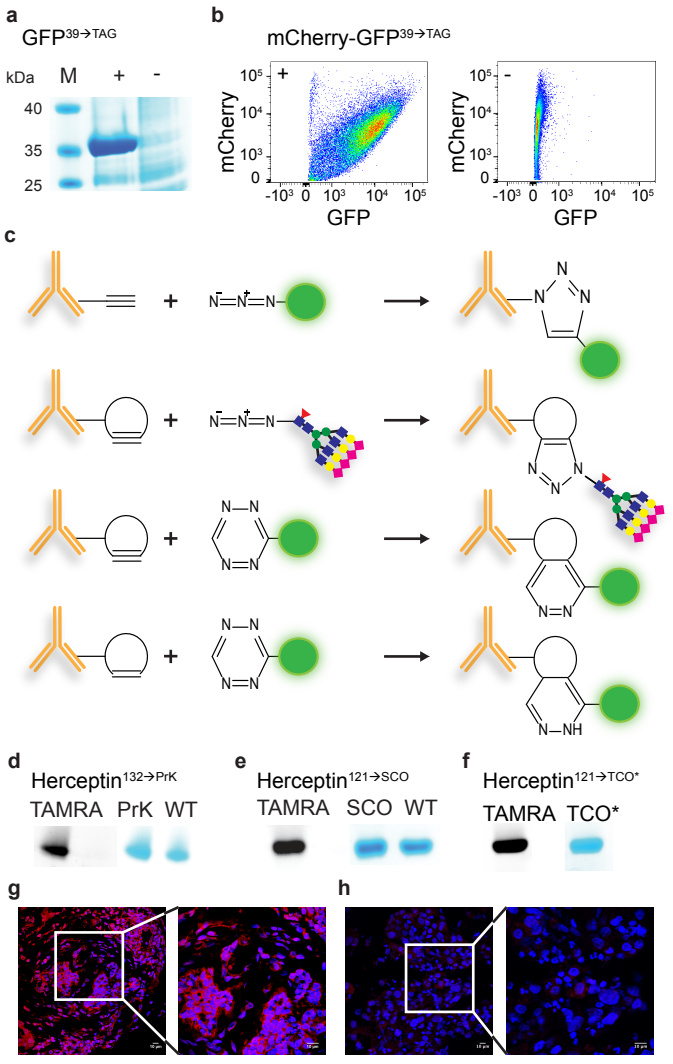
The European Institute of Oncology (IEO) Division of Biostatistics selected from its institutional database consecutive breast cancer (BC) patients fulfilling the following criteria: i) histologically proven invasive BC treated by neoadjuvant therapy; ii) any age (pre- or postmenopausal status allowed); iii) any intrinsic subtype (Luminal A/B-like, Her-2 positive, Triple Negative subtypes allowed); All the patients prospectively entered the IEO BC database and were discussed at the weekly multidisciplinary meeting. Data on patients' medical history, concurrent diseases, surgery, pathological evaluation, radiotherapy, neoadjuvant systemic treatments, and clinico-pathological results of pre- and post-neoadjuvant treatment staging procedures were retrieved. All the biopsies were fixed in 4% buffered formalin for less than 24 hours immediately after the core biopsy procedure. All the surgical samples were fresh sampled in accordance to the criteria issued by Provenzano et al. (2015) [37] and fixed in 4% buffered formalin for less than 24 hours. All the biopsies and surgical samples were routinely processed and embedded in paraffin. Detailed information regarding tumor type and grade, ER/PgR and Her-2 *status*, and Ki-67 labeling index were available in all the cases. ER/PgR and HER2 immunoreactivity was assessed in line with the clinical practice procedures applicable at diagnosis. Her-2 immunoreactivity was assessed using

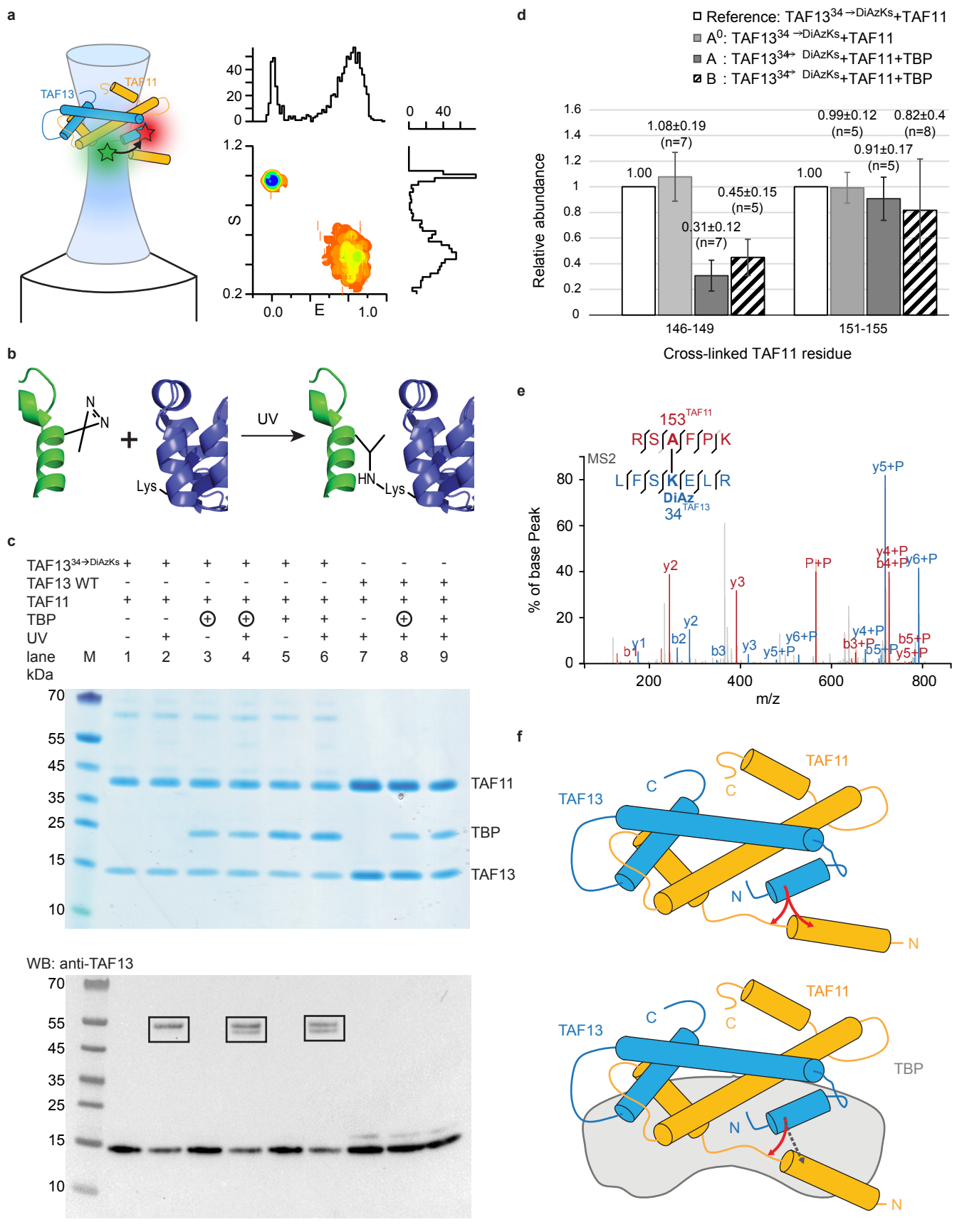
the monoclonal antibody CB11 (Novocastra, 1:800) from 1995 till 2005, and the HercepTest (Dako) thereafter. Cases classified as Her-2 2+ by immunohistochemistry were tested by FISH analysis with Vysis probes, in accordance with the ASCO/CAP guidelines [38]. Ki-67 labeling index was assessed by the Mib-1 monoclonal antibody (Dako, 1:200), by counting at least 500 invasive tumor cells, independent of their staining intensity and without focusing on hot-spots [39]. Tumors were classified as Luminal A-like (ER and PgR positive, absence of Her-2 overexpression and Ki-67 <20%), Luminal B-like (ER positive, Her-2 negative and at least one of Ki-67 ≥20% and PgR <20%), Luminal B-like/Her-2 positive (ER and Her-2 positive, any PgR and Ki-67), Her-2 positive (Her-2 3+ and/or amplified by FISH, ER/PgR negative) and Triple Negative (ER, PgR and Her-2 negative) in accordance with St. Gallen recommendations [40]. For tumor specific information please refer to **Supplementary Table 3**. All the patients included gave an informed consent for using their clinico-pathological data and samples for research purposes at the time of admission to the hospital, and the study was approved by the IEO Review Board.

Online Methods references

23. Simpson, J.T. & Durbin, R. Efficient de novo assembly of large genomes using compressed data structures. *Genome research* **22**, 549-556 (2012).
24. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
25. Magoc, T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963 (2011).
26. Berger, I., Fitzgerald, D.J. & Richmond, T.J. Baculovirus expression system for heterologous multiprotein complexes. *Nature biotechnology* **22**, 1583-1587 (2004).
27. Nie, Y., Bieniossek, C. & Berger, I. ACEMBL Expression System, User Manual. **Vers. 09.11** (2009).
28. Milles, S. & Lemke, E.A. Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153. *Biophysical journal* **101**, 1710-1719 (2011).
29. Milles, S. et al. Click strategies for single-molecule protein fluorescence. *Journal of the American Chemical Society* **134**, 5187-5195 (2012).
30. Maiolica, A. et al. Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Molecular & cellular proteomics : MCP* **6**, 2200-2211 (2007).
31. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical chemistry* **75**, 663-670 (2003).
32. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372 (2008).
33. Tomko, R.J., Jr. et al. A Single alpha Helix Drives Extensive Remodeling of the Proteasome Lid and Completion of Regulatory Particle Assembly. *Cell* **163**, 432-444 (2015).
34. Chen, Z.A., Fischer, L., Cox, J. & Rappsilber, J. Quantitative Cross-linking/Mass Spectrometry Using Isotope-labeled Cross-linkers and MaxQuant. *Molecular & cellular proteomics : MCP* **15**, 2769-2778 (2016).
35. Thakur, S.S. et al. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Molecular & cellular proteomics : MCP* **10**, M110 003699 (2011).
36. Tyagi, S. & Lemke, E.A. Genetically encoded click chemistry for single-molecule FRET of proteins. *Methods in cell biology* **113**, 169-187 (2013).
37. Provenzano, E. et al. Standardization of pathologic evaluation and reporting of postneoadjuvant specimens in clinical trials of breast cancer: recommendations from an international working group. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **28**, 1185-1201 (2015).
38. Wolff, A.C. et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31**, 3997-4013 (2013).
39. Polley, M.Y. et al. An international Ki67 reproducibility study. *Journal of the National Cancer Institute* **105**, 1897-1906 (2013).
40. Goldhirsch, A. et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **24**, 2206-2223 (2013).


MultiBacTAG
PylRS
tRNA
PylRS
tRNA
Tn7 transposition
transfection
transfection
X-ray phase problem
custom biologics
cross-linking
Lys
protein complexes
tandem
recombineering
Cre
X
pIDC
pIDK
pIDS
pACEBac1,2
1
PrK
2
SCO
3
BOC
4
TCO*
5
BCN
6
DiAzKs
smFRET
1.2
0.2
S
0.0
E
1.0


a
GFP39→TAG
kDa
M
+
-
40
35
25
b
mCherry-GFP39→TAG
+
-
mCherry
GFP
c
d
Herceptin132→PrK
TAMRA
PrK
WT
e
Herceptin121→SCO
TAMRA
SCO
WT
f
Herceptin121→TCO*
TAMRA
TCO*
g
h


a
TAF13
TAF11
40
20
0
1.2
S
0.2
0.0
E
1.0
0
40
b
N
N
+
Lys
UV
HN
Lys
c
TAF13^34→DiAzKs + + + + + + - - -
TAF13 WT - - - - - - + + +
TAF11 + + + + + + + + +
TBP - - ⊕ ⊕ + + - ⊕ +
UV - + - + - + + + +
lane M 1 2 3 4 5 6 7 8 9
kDa
70
55
45
35
25
15
10
TAF11
TBP
TAF13
WB: anti-TAF13
70
55
45
35
25
15
10
d
Reference: TAF13^34→DiAzKs+TAF11
A^0: TAF13^34→DiAzKs+TAF11
A : TAF13^34→DiAzKs+TAF11+TBP
B : TAF13^34→DiAzKs+TAF11+TBP
Relative abundance
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1.00
1.08±0.19 (n=7)
0.31±0.12 (n=7)
0.45±0.15 (n=5)
1.00
0.99±0.12 (n=5)
0.91±0.17 (n=5)
0.82±0.4 (n=8)
146-149
151-155
Cross-linked TAF11 residue
e
153^TAF11
R S A F P K
L F S K E L R
DiAz
34^TAF13
MS2
% of base Peak
80
60
40
20
0
200
400
600
800
m/z
y5+P
y4+P
b4+P
y6+P
P+P
y2
y3
y2
b2
y1
b1
b3
y3
y6+P
y5+P
y4+P
b3+P
b5+P
b5+P
y5+P
f
TAF11
TAF13
C
C
N
N
TAF11
TAF13
C
C
TBP
N
N